

Noise Spectrum Subtraction Method and System

FIELD OF THE INVENTION

This invention is in the field of noise subtraction techniques, and relates to a noise spectrum subtraction method and a voice-processing unit utilizing the same for use in a voice operated system.

5 BACKGROUND OF THE INVENTION

Voice operated systems are typically utilized in communication devices, such as phone devices and computers, as well as in toys. These systems typically comprise such main constructional components as an A/D converter for receiving an input analog voice signal, a vocoder, an operating system, a communication interface associated with an output port, and a voice recognizer (typically implemented as a separate DSP chip).

During a transmission operational mode of the communication device (e.g., mobile phone), the input analog voice signals (e.g., generated by a microphone) are digitized by the converter. In the conventional devices, the digitized voice signals are supplied to the vocoder for compression of the voice samples to reduce the amount of data to be transmitted through the interface unit to another communication device (e.g., mobile phone), and are concurrently supplied to the voice recognizer. The latter receives the digitized voice samples as input, parameterizes the voice signal and matches the parameterized input signal to reference voice signals. The voice recognizer typically either provides the identification of the matched signal to the operating system, or, if a phone number is associated with the matched signal, provides the associated phone number.

A technique utilizing the application of a voice recognition function to a compressed digitized signal has been developed and disclosed in US Patent No. 6,003,004 assigned to the assignee of the present application.

It is a well-known problem of voice operated systems that background noise added to speech can degrade the performance of digital voice processors used for speech compression, recognition, authentication, etc. Thus, to improve the quality of voice recognition, it is necessary to reduce the background noise in a speech signal.

Various noise reduction techniques have been developed and disclosed, for example, in the article S.F. Boll "Suppression of Acoustic Noise in Speech Using Spectral Subtraction", IEEE Transactions in Acoustics, Speech and Signal processing, 1979, V. 27, N. 2, pp. 113-120. According to the known techniques, the noise suppression of the digital signal is typically carried out before the signal is supplied to the vocoder (i.e., prior to signal compression). This approach is therefore computationally intensive and slow. This is a serious drawback when dealing with mobile phones, since the processing requirements of noise suppression and voice recognition pose a severe processing load on the mobile phone and may obstruct its operation. It is known to use an additional DSP chip for noise suppression.

20 SUMMARY OF THE INVENTION

There is therefore a need in the art to facilitate noise reduction in voice operated systems by providing a novel noise spectrum subtraction method and a voice processing unit utilizing the same.

The main idea of the present invention consists of applying a noise reduction to a digital signal representative of a voice signal, after the digital signal being compressed. This simplifies the computation.

There is thus provided according to one aspect of the present invention, a method for reducing noise in a voice signal, the method comprising the steps of:

- (i) processing a compressed digital signal representative of the voice signal including a speech component and a noise component; and
- (ii) determining the noise component to be subtracted from the compressed digital signal.

5 In a preferred embodiment of the invention, the compressed digital signal is based on a set of linear prediction coding (LPC) coefficients and a residual signal, and is obtained by applying LPC analysis to the voice signal. To this end, a digital signal may be divided into a series of frames representative of the voice signal including a speech component and a noise component to be subtracted. The frame
10 may, for example, represent about 20msec of the digital signal. Preferably, the frame is composed of M digitized speech samples, and the set of LPC coefficients contains p coefficients, such that the ratio p/M is in the range of 0.1-0.25. LPC analysis is applied to all frames, thereby obtaining the compressed digital signal representative of the voice signal.

15 Preferably, the processing of the compressed digital signal is based on the following: determination of a power spectrum of the noise component during a non-speech activity and calculation of its average value, calculation of a power spectrum estimator of the compressed digital signal with a reduced noise component, determination of an autocorrelation function of this signal, and
20 determination of modified LPC coefficients. The modified LPC coefficients represent the speech component with the reduced noise spectrum. To determine the noise spectrum, a calculation involving a Fourier transform can be applied to the compressed digital signal. To determine the autocorrelation function of the compressed digital signal with the reduced noise component, an inverse Fourier
25 transform may be applied to the estimated power spectrum of the signal with the reduced noise component.

According to another aspect of the present invention, there is provided a voice processing unit for use in a voice operated system, the voice processing unit comprising a noise reduction utility interconnected between a voice coding utility
30 and a voice recognition utility, the noise reduction utility being operable for

processing a compressed digital signal representative of an input voice signal received from the voice coding utility and generating an output compressed digital signal with reduced noise spectrum.

According to yet another aspect of the present invention, there is provided a
5 voice operated system comprising an input port for receiving an input voice signal, an analog-to-digital converter for processing the input signal to generate a digital output indicative thereof, a voice processing utility for processing the digital signal and generating a compressed digital signal representative of the input voice signal, a voice processing unit, a system interface utility, and a control module, which is
10 interconnected between the voice processing utility and the voice processing unit, and is connected to the system interface to operate it in response to a speech signal, the voice processing unit comprising:

- 15 - a noise reduction utility coupled to the voice processing utility and operable to process said compressed digital signal and generate an output compressed digital signal with reduced noise spectrum; and
- a voice recognition utility coupled to the noise reduction utility for processing said output compressed digital signal with reduced noise spectrum.

BRIEF DESCRIPTION OF THE DRAWINGS

20 In order to understand the invention and to see how it may be carried out in practice, a preferred embodiment will now be described, by way of non-limiting example only, with reference to the accompanying drawings, in which:

Fig. 1 is a block diagram of a voice operated system according to the invention; and

25 Fig. 2 is a flow chart of main operational steps of a voice processing unit of the system of Fig. 1.

DETAILED DESCRIPTION OF THE INVENTION

Referring to Fig. 1, there are illustrated the main components of a voice operated system 10 according to the invention (e.g., a mobile phone device). These components include the following: an A/D converter 14 for receiving an analog voice signal coming from an input port 12 (e.g., a microphone), a system interface utility 20 associated with an output port (not shown), a voice processing utility (vocoder) 22, a voice processing unit 24, and a control unit (module) 26, which is interconnected between the vocoder 22 and the voice processing unit 24, and is connected to the system interface utility 20. The voice processing unit 24 comprises a noise reduction utility 28 coupled to the vocoder 22 through the control unit 26, and a voice recognition utility 29 coupled to the noise reduction utility 28.

The operation of the system 10 will now be described with reference to Fig. 2. Initially, the A/D converter 18 converts the input analog voice signal into an output digital signal, and supplies the digital output to the vocoder 22 (step 30).
The vocoder 22 is operable by suitable software to compress the digital signal.

In the present example, a voice compression algorithm based on LPC analysis is utilized. It should, however, be noted that any other suitable technique can be used for digital signal compression, for example, the voice quantization technique.

Thus, in the present example, to compress the input digital signal, it is divided into a series of frames (step 32). Each frame contains M samples $x(m)$, wherein $m=1,2,3,\dots,M$, and typically represents 20msec of the input signal.

The signal $x(m)$ is typically a sum of a speech signal component, $s(m)$, and a stationary additive background noise component, $n(m)$, which is to be reduced, that is:

$$x(m)=s(m)+n(m) \quad (1)$$

The vocoder performs LPC analysis on each frame and provides an output compressed signal thereof (step 34). Generally, the LPC analysis can be applied to at least some samples of at least one frame.

As a result, the given signal sample $x(m)$ is represented in the following form:

$$x(m) = \sum_{i=1}^p a_i x(m-i) + \varepsilon(m) = \sum_{i=1}^p a_i [s(m-i) + n(m-i)] + \varepsilon(m) \quad (2)$$

- 5 wherein a_i are the LPC coefficients and $\varepsilon(m)$ is a residual signal, all being the parameters of the frame. Each frame has p LPC coefficients a_i .

The vocoder further parameterizes the residual signal $\varepsilon(m)$ in terms of at least pitch and gain values (step 36).

The above coding scheme usually results in a compression factor of
10 approximately 8-11. The output of the vocoder 22 is supplied to the noise reduction utility 26 through the control module 26. The noise reduction utility is operable to determine a power spectrum of the noise component during a non-speech activity (step 38), and to remove the power spectrum of the noise component from the noisy speech signal. In the present example, the power
15 spectrum of a signal $x(m)$ is denoted by $|X(\omega_m)|^2$ and is calculated as follows:

$$X(\omega_m) = S(\omega_m) + N(\omega_m) = H(\omega_m) \cdot E(\omega_m)$$

$$H(\omega_m) = \frac{1}{1 + \sum_{i=1}^p a_i \cdot e^{-j\omega_m k}} \quad (3)$$

$$|X(\omega_m)|^2 = |H(\omega_m)|^2 |E(\omega_m)|^2$$

20

wherein $S(\omega_m)$, $N(\omega_m)$ and $E(\omega_m)$ are Fourier transforms of $s(m)$, $n(m)$ and $\varepsilon(m)$, respectively. It should be noted that, for non-speech frames, $X(\omega_m) = N(\omega_m)$.

In the present invention, it is assumed that the power spectrum of $\varepsilon(m)$ is constant, i.e., $|E(\omega_m)|^2 = E_0^2$. By using Parseval theorem, the value of E_0^2 can be
25 estimated as follows:

$$E_0^2 = \frac{1}{M} \sum_{m=1}^M |E(\omega_m)|^2 = \frac{1}{M} \sum_{m=1}^M \varepsilon(m)^2 \quad (4)$$

The noise reduction utility determines the noise power spectrum $|N(\omega_m)|^2$ during the non-speech activity and calculates its average value $\langle |N(\omega_m)|^2 \rangle$ over 5 non-speech frames (step 40), as follows:

$$\langle |N(\omega_m)|^2 \rangle = \mu(\omega_m) \quad (5)$$

Using the above expressions, the noise reduction utility 28 determines the 10 speech signal power spectrum estimator $\hat{S}(\omega_m)$ with reduced noise component (step 42), as follows:

$$\hat{S}(\omega_m) = |H(\omega_m)|^2 \cdot E_0^2 - \mu(\omega_m) \quad (6)$$

In equation (6), all the $\hat{S}(\omega_m)$ samples which are less than zero are replaced by zeros (clipping condition). It should be noted that $\hat{S}(\omega_m)$ is advantageously based only on p LPC coefficients a_i ($p \ll M$) and on the total energy of the residual signal.

As known, for example, from the disclosure in the following book: A.V. 20 Oppenheim et al., "Digital Signal Processing", Prentice Hall, Inc., Englewood Cliffs, NJ, 1975, p. 557, the inverse Fourier transform of $\hat{S}(\omega_m)$ is the autocorrelation function $r(n)$ of the signal, that reads:

$$r(n) = \frac{1}{M} \sum_{m=1}^M \hat{S}(\omega_m) \cdot e^{j\omega_m n} = \sum_{m=1}^M s(m) \cdot s(m-n) \quad (7)$$

25

Based on the above equation, the noise reduction utility 28 determines modified LPC coefficients \hat{a}_k (step 44). To implement this, any known suitable

technique can be used, for example, those disclosed in the book: Rabiner et al., "Fundamentals of Speech Recognition", Prentice Hall, 1993, pp 97-121. The modified LPC coefficients \hat{a}_k represent the compressed digital signal with the reduced noise component.

5 Thus, the noise recognition utility determines the modified LPC coefficients, generates an output compressed digital signal indicative thereof, and supplies this signal to the voice recognition utility 29, which utilizes the same for performing the voice recognition.

10 It should be noted that the noise reduction utility 28 can also produce various LPC based parameters, such as cepstrum coefficients, MEL cepstrum coefficients, line spectral pairs (LSPs), reflection coefficients, log area ratio (LAR) coefficients, and the like.

15 Those skilled in the art will readily appreciate that various modifications and changes can be applied to the preferred embodiment of the invention as hereinbefore exemplified without departing from its scope defined in and by the appended claims. For example, any suitable technique can be used to determine modified LPC coefficients. The voice operated system utilizing the voice processing unit according to the invention may be of any suitable type, other than the mobile phone device described above.

SEARCHED
INDEXED
SERIALIZED
FILED